Available online at www.sciencedirect.com

ScienceDirect

Procedia Engineering 89 (2014) 1023 – 1030

Procedia Engineering

www.elsevier.com/locate/procedia

16th Conference on Water Distribution System Analysis, WDSA 2014

Urban Water Demand Forecasting for the Island of Skiathos

D. Kofinas[a,b], N. Mellios[a,b], E. Papageorgiou[a,c], C. Laspidou[a,b,]*

[a]Information Technologies Institute, Center for Research & Technology Hellas-CERTH, 6[th] km Charilaou-Thermi Rd., Thermi 57001, Greece
[b]Department of Civil Engineering, University of Thessaly, Pedion Areos, Volos 38334, Greece
[c]Department of Computer Engineering, Technological Educational Institute of Central Greece, 35100 Lamia, Greece

Abstract

We present an analysis of historical water demand data from the utility of Skiathos, Greece and demonstrate suitable demand forecasting methodologies. We apply linear and nonlinear forecasting methods to a three-year time series water demand. The best fit for quarterly averaged data was observed for the Winters' additive method; for monthly-averaged data, ARIMA, Artificial Neural Network and a hybrid approach performed best. Given the intense seasonality of demand in Skiathos, monthly time series proved to be the best data set for forecasting, while the best forecasting method was the hybrid, which combines the advantages of ARIMA and Artificial Neural Networks.

Peer-review under responsibility of the Organizing Committee of WDSA 2014

Keywords: water distribution networks; urban water demand; forecasting; time series; ARIMA; artificial neural networks

1. Introduction

Greek islands are mainly inhabited by small communities with inadequate and ageing water distribution networks. Many of these small communities face a number of specific issues; namely they: (i) represent the core of Greece's tourist industry; (ii) experience a heavy seasonal rise in population; (iii) mostly discharge their treated effluents to the sea and other sensitive ecosystems; and (iv) deal with issues of insularism [1]. Tourism results in high seasonal water demand variability, with summer water demand being over twice as high as winter demand and most of it covering the needs of large hotel units and filling-up swimming pools. The island of Skiathos has a small water distribution network with a total of about 3,500 water meters; its water demand presents high variability and seasonality, while its

* Corresponding author. Tel.: +30-2421074147; fax: +30-2421074165.
E-mail address: laspidou@iti.gr

Peer-review under responsibility of the Organizing Committee of WDSA 2014
doi:10.1016/j.proeng.2014.11.220

network has reportedly some of the highest water leakages in Greece—as high as 50% or more. The island faces serious water shortage issues, with aquifer salinization and deteriorating groundwater quality; since all urban water needs are covered by groundwater, it is important that network leakage is reduced to a minimum. Any significant improvement in leakage prevention would involve a series of strategic decisions and actions on behalf of the local water utility. At the bottom of these actions lies a reliable water demand forecasting routine that will be based on historical data and will be able to capture its stochastic nature. Thus, urban water demand forecasting is a key step for water utilities in order to optimize water network operation.

Water demand is an important parameter that needs to be taken under consideration in the operation of water supply and distribution systems. The operation of these systems requires frequent adjustments in response to variations in demand in order to optimize its operation and minimize distribution costs. Water demand includes household consumption, but also network leakage, since it is the combined amount that is put into supply [2]. Water demand forecasting also plays a significant role in managing and planning water supply operations and water conservation strategies. Short-term demand forecasting is needed in order to increase the stability of urban freshwater supply by adjusting water supply to actual demand and consumption, thus resulting in the optimal and timely use of water resources. Demand forecasting is also critical for optimal pump scheduling and thus, for supplying water in a more energetically efficient manner [2, 3].

In recent years, different methods and approaches to urban water demand forecasting have been proposed. These methods vary from the simplest historical extrapolation to sophisticated analytical models; therefore the choice of an appropriate model depends on the purpose of the forecasting required by water utility companies, and the quality and quantity of data [4]. Recent forecasting methods have been developed mainly to control water distribution operation systems nearly in real time, based on time series data that are collected sequentially over various time periods. Several time series forecasting models that forecast future water demand on the basis of past observations and associated error terms are found in the literature [3]. In general, the time series forecasting models of water demand can be divided into deterministic and stochastic models. Deterministic models are used to model time series by a series of seasonal, weekly and daily patterns considering physical nature. Stochastic models usually adopt a numerical approach since they are usually formulated by using statistical and probabilistic models that are built on historical data [5, 6].

In water demand time series modeling, common stochastic models are: the pure random (or white noise) model, the autoregressive (AR) model, the moving average (MA) model, the autoregressive moving average (ARMA) model, the autoregressive integrated moving average (ARIMA) model and the seasonal autoregressive integrated moving average (SARIMA) models [5, 6, 7]. These models, known as traditional statistical models, are linear in that predictions of the future values are constrained to be linear functions of past observations. Because of their relative simplicity in understanding and implementation, linear models have been the main research focus and tools applied extensively during the past few decades [8].

Although ARIMA models are quite flexible in that they can represent several different types of time series, i.e., pure AR, pure MA and combined AR and MA (ARMA) series, their major limitation is the pre-assumed linear form of the model. That is, a linear correlation structure is assumed among the time series values and therefore, no nonlinear patterns can be captured by the ARIMA model. To overcome the restriction of ARIMA model and to account for certain nonlinear patterns observed in real problems, artificial neural networks (ANNs) have been proposed in the literature. ANNs have been suggested as an alternative to time series forecasting to deal with linear and nonlinear relationships. The major advantage of NNs is their flexible nonlinear modeling capability. With ANNs, there is no need to specify a particular model form. Rather, the model is adaptively formed based on the features presented from the data. This data-driven approach is suitable for many empirical data sets where no theoretical guidance is available to suggest an appropriate data generating process.

A recent study reviews the literature on urban water demand forecasting published from 2000 to 2010, in order to identify methods and models useful for specific water utility decision making problems [9]. It presents an annotated reference list of the methods and models for water demand forecasting identifying the forecast variable and periodicity, the determinants used and the forecast horizon. Results show that although a wide variety of methods and models have attracted attention, applications of these models differ, depending on the forecast variable, its periodicity and the forecast horizon. The popular models used for short-term forecasting are ARIMA and ANNs. They gained popularity due to the advantageous characteristics of each one such as the ability to catch the general trend and seasonal fluctuation, for the ARIMA, and the ability to catch nonlinear components requiring less statistical training for the ANNs.

Also, hybrid models have been suggested, combining the ARIMA model and neural networks, in order to overcome the deficiencies of single models. Su et al. [10] used a hybrid model to predict a time series of reliability data with growth trend. Their results showed that the hybrid model produced better predictions than either the ARIMA model or the neural network by itself. Zhang [11] proposed a hybrid ARIMA and ANN model to take advantage of the two techniques and applied the proposed hybrid model to some real data sets. He concluded that the combined model can be an effective way to improving predictions achieved by either of the models used separately. Faruk [12] proposed a hybrid neural network and ARIMA model, developed for the Buyuk Menderes river, for water quality time series prediction. He indicated that the approach of combining the strengths of the conventional and ANN techniques provides a robust modeling framework capable of capturing the nonlinear nature of the complex water quality time series, thus producing more accurate forecasts.

In general, research activities in water demand forecasting with ANNs suggest that ANNs can be a promising alternative to the traditional linear methods. In Jentgen et al. [13], authors exploited the performance of ARIMA models and ANNs for forecasting, which are often compared—with mixed conclusions—in terms of the superiority in forecasting performance. The ARIMA model cannot deal with nonlinear relationships while the neural network model alone is not able to handle both linear and nonlinear patterns equally well. Thus, hybrid models were investigated that are capable of exploiting the strengths of traditional time series approaches and ANNs [12]. Due to the present complexity in real-life time series efficient approaches are needed.

The goal of this study is to evaluate methodologies and identify the most suitable method for forecasting urban water demand relevant to water-utility decision-making problems. These methods will be applied to three-year historical water demand data of Skiathos, Greece for short-term urban water demand forecasting. We investigate the use of the two most popular time series forecasting methods for water demand, ARIMA and ANN and extend our analysis to other methods, such as the Winters' Additive Exponential Smoothing and hybrid ARIMA-ANN to obtain more reliable and accurate short-term forecasting. Based on the results, we conclude on the most suitable method for short-term urban water demand prediction for Skiathos and evaluate the accuracy of forecasted values.

2. Materials and Methods

The period on which the models are calibrated and validated is approximately three years, from January 2011 to November 2013. For this time period, in collaboration with the Skiathos Municipal Enterprise for Water Supply and Sewerage (www.deyaskiathos.gr), we obtained daily groundwater pumping values. From the daily values, we produced monthly and quarterly averages, so as to investigate the trends and seasonality of water demand in micro and macro scale. The time series' values do not take into account network leakage, since they represent water measurement at the network inlet. Although this does not correspond to actual water consumption, it represents the water demand exerted on local water resources, so it is a valid water demand data set. Before applying the various methods on the data (presented below), the daily time series is normalized, after the outliers are removed. Furthermore, these values refer to the amounts of water, pumped each day and stored into a tank. If one day's demand is less than the amount of water pumped at that day, the surplus left in the tank is used to cover next day's demand additional to what will be pumped. This means that this time series is not quite representative of the daily fluctuations of demand, but is adequately accurate for the monthly averaged scale. Therefore, our analysis shows that the models are capable of capturing the intense seasonality of data, when monthly and quarterly averaged time series are used. The forecasting models that we chose to apply are both linear and nonlinear: specifically, they are the seasonal ARIMA, the Winters' Additive Exponential Smoothing, ANN and a hybrid approach, as described below.

2.1. ARIMA

Autoregressive integrated moving average (ARIMA) is one of the most important and widely used linear models in time series forecasting during the past three decades [14]. The popularity of the ARIMA model is due to its statistical properties as well as the well-known Box–Jenkins methodology in the model building process. In addition, various exponential smoothing models can be implemented by ARIMA models [15]. In an ARIMA model, the future value of a variable is assumed to be a linear function of several past observations and random errors. The linear function is based upon three parametric linear components: AR, integration (I), and MA method [16,17].

A SARIMA model is expressed as ARIMA (p, d, q) × (P, D, Q), where (p, d, q) is the non-seasonal part of the model and (P, D, Q) is the seasonal part of the model [18]. The value of p is the order of non-seasonal AR and is determined from the partial autocorrelations of the appropriately differenced time series. If the partial autocorrelations cut off after a few lags, the last lag with a large value would be the estimated value of p. If the partial autocorrelations do not cut off, we either have a moving average model ($p=0$), or an ARIMA model with positive p and q. The value of d is the number of regular differencing and is estimated by considering the autocorrelation plots. When the autocorrelations die out quickly, the appropriate value of d has been found. The value of q is the order of non-seasonal MA. It is found from the autocorrelations of the appropriately differenced series. If the autocorrelations cut off after a few lags, the last lag with a large value would be the estimated value of q. If the autocorrelations do not cut off, we either have an autoregressive model ($q=0$) or an ARIMA model with a positive p and q. Respectively, P is the number of seasonal AR terms, D is the number of seasonal differences and Q is the number of seasonal MA. Identification of a seasonal series is much more difficult. Box-Jenkins describe methods for model identification, but the user must be very skilled and experienced to successfully identify the model order. Usually, trial and error must be used. It is preferable to keep the number of parameters to a minimum, so the values of p, P, q, Q, d, and D that are selected should be less than or equal to two [19].

The ARIMA model order is identified by the trial and error method with use of IBM SPSS Statistics 20. The criteria are the optimal combination of 4 statistical amounts: the R square, the Root Mean Square Error (RMSE), the Mean Absolute Percentage Error (MAPE) and the Mean Absolute Error (MAE). The R square should be maximized, approaching the value of 1. On the contrary, RMSE, MAPE and MAE take values depending on the order of magnitude of the variable that is being forecasted and should be minimized, ideally nulled.

2.2. Winters' Additive Exponential Smoothing

The Winters' Additive method [20] is applicable when the time series contains a seasonal component. This method assumes that the time series is composed by a linear trend and a seasonal cycle, it constructs three statistically correlated series (smoothed, trend and seasonal) and projects forward the identified trend and seasonality. The additive method is preferred when the seasonal variations are roughly constant through the series. With the additive method, the seasonal component is expressed in absolute terms in the scale of the observed series, and in the level equation the series is seasonally adjusted by subtracting the seasonal component. Within each year, the seasonal component will add up to approximately zero. Therefore, the Winters' additive is based on the implicit assumption that the different components affect the time series, in an additive fashion, as shown below [19]:

$$\textit{Forecast} = \textit{Level} + \textit{Trend} + \textit{Seasonal} \tag{1}$$

2.3. Artificial Neural Networks

An ANN model is a massively parallel-distributed processor that has a natural propensity for storing experiential knowledge and making it available for later use [21]. It resembles the brain in two respects. The ANN models can recognize trends, patterns, and learn from their interactions with the environment. The most extensively studied and used ANN models are the multilayer feed forward networks [22], which allow information transfer only from an earlier layer to the next consecutive layers. Hence, the ANN model performs a nonlinear functional mapping from the past observations to the future value y_t. Thus, the NN is equivalent to a nonlinear AR model. A neural network must be "trained" to determine the values of the weights that will produce the correct outputs. In a training step, a set of input data is used for training and is presented to the network many times. The performance of the network is tested after the training step is stopped. The back-propagation algorithm adjusts the weights in the steepest descent direction (negative of the gradient) [23]. The ANN model is applied with the Zaitun time series software.

2.4. Hybrid approach

The combination of the ARIMA and ANN models is performed to use each model capability to capture different patterns in the data. An ARIMA model is not sufficient if there are still nonlinear correlation structures left in the residuals. Therefore, the residuals can be modeled by using ANNs to discover nonlinear relationships. The methodology consists of two steps: (1) in the first step, the ARIMA model is developed to forecast water demand; and (2) in the second step, the ANN model presented above is used to simulate the residuals from the ARIMA model. The hybrid model is built using the IBM SPSS Statistics 20 and Zaitun Time Series software packages.

3. Results and Discussion

Fig. 1 shows a plot of the quarterly water demand throughout the three years that we had available data and the forecasted water demand for the following year. Water demand is plotted as a unitless normalized value, in order to make relative values more clear. Through the quarterly averaged data processing, in which we have reduced daily values for three years to four values per year, it emerges that there is a peak in summer, almost six times higher than the winter demand, while spring and autumn periods are ascending and descending, respectively. The dramatic summer increase is due to the intense summer touristic activity—Skiathos island is supposedly one of the most touristic islands in Greece—while temperature increase plays a role as well. It should be noted here that leakage is not constant throughout the year, but it is related to pressure in the water distribution network and therefore to demand. The distribution of water demand is seasonal and faintly follows an ascending trend through the years. The ascending trend could be more justified if longer time series data were available; however, it corresponds to the observed annual population increase.

In Fig. 1 and 2, in which diagrams of normalized water demand through time are presented, the blue- shaded area is the estimation period—the period for which we have historical data and on which we are basing our forecast—while the green-shaded area is the forecast period for year 2014. In Fig. 1, we see that the method uses first-year data to be "trained" and is only capable of predicting values in subsequent years. The more "training" the model has, the better predictions it can produce. In our case, three years is a very limited time, so this analysis shows how good forecasts can be even with the limited data set of only three years. Having attempted to obtain demand forecast values using all models presented herein, we have concluded that the Winters' Additive Exponential Smoothing is the only model that is capable of producing a forecast based on such few values, so we only show these results.

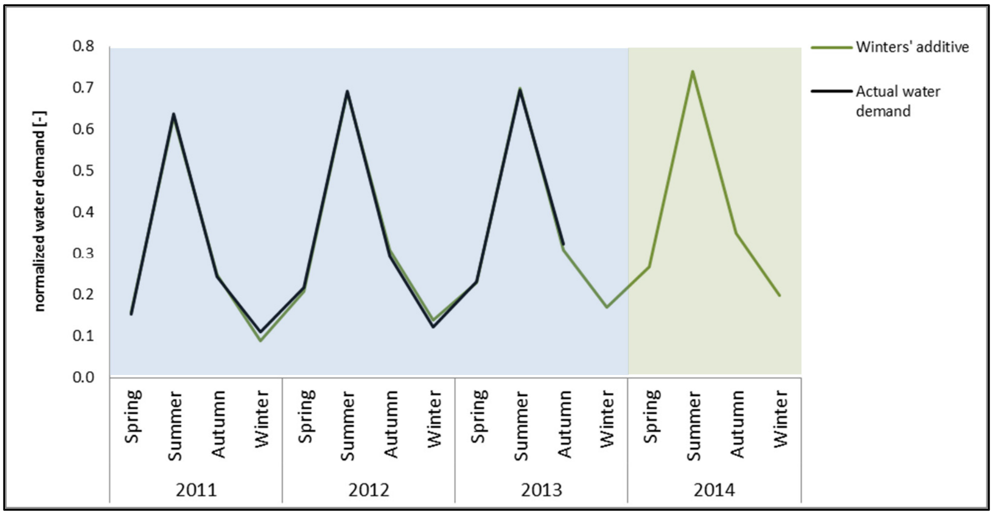

Fig. 1 Fit and forecast of the normalized quarterly water demand.

Table 1. Statistical amounts for the adequacy of the Winters' Additive Smoothing Exponential model for the quarterly water demand

R square	RMSE	MAPE	MAE
0.997	0.014	5.009	0.010

Table 1 presents the corresponding statistics for the Winter's model that confirm the fact that the predicted values correspond quite well to the data, through the high value of R square that is almost 1.

The RMSE, MAPE and MAE are relatively low. The areas with the poorest fitting concern the winter period.

We repeat the same analysis for the monthly average water demand. In this case, several models give quite good predictions, as shown in Fig. 2. Observing the monthly averaged water demand, it can be deduced that the spring ascending slopes are smoother than the autumn descending, which is related to the touristic distribution. Water demand distribution during the winter months is more variable, possibly due to the more intense weather variability, causing the models' relative difficulty to fit. Although the ANN seems to fit best the variability throughout the years, it does not seem capable to capture the generally ascending trend as the linear models do—namely, ARIMA (2,0,2) × (1,1,0) and Winters Additive Exponential Smoothing. The hybrid seems to fit almost as well as the ANN; furthermore, it captures the general ascending trend from one year to the next that linear models can capture as well.

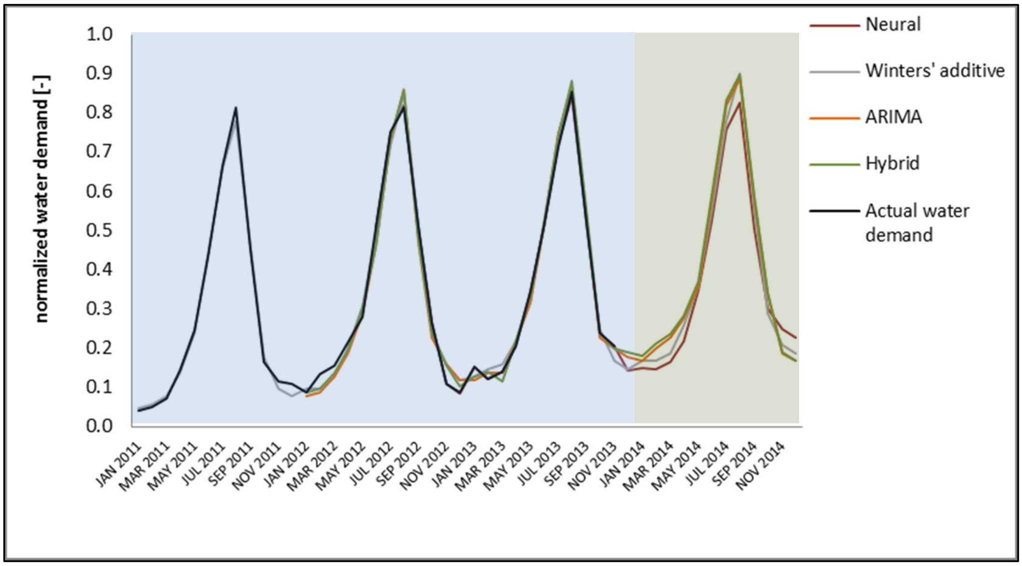

Fig. 2. Fit and forecast of the normalized monthly water demand

Table 2. Statistical amounts for the adequacy of 4 models for the monthly water demand

Model	R square	RMSE	MAPE	MAE
Neural	0.999	0.053	0.649	0.003
ARIMA	0.987	0.158	10.493	0.025
Winters'Additive	0.940	0.149	10.797	0.022
Hybrid	0.990	0.125	6.323	0.016

In Table 2, in which we present all relevant statistics of the forecast methods used, one may assume that ANN is the better fitting model, with high R square and approximately nulled RMSE, MAPE and MAE. To show the same analysis in a different format, we present scatter plots of observed and predicted monthly water demand values for the four different methodologies presented herein. Observed values are plotted on the y-axis, while simulated values are shown on the x-axis. A perfect fit would have all data falling on a 45-degree line that goes through the origin. We see that all models have a very good fit, with the neural network method appearing to have almost all points on the 45-degree line (Fig. 3). However, the statistical amounts in such few periods and the scatter plots are not the only criteria to be taken into account. By simply observing the forecast in Fig. 2, we see that the linear models, and therefore the hybrid, can capture the ascending trend. Thus, we conclude that the hybrid model is a powerful forecasting tool for water demand, even when a limited data set is only available, while the ANN can provide really good fits as well. Naturally, the results will become stronger, if more extensive data sets become available.

In an effort to process the daily water demand, the data is divided into three periods per year, the ascending and descending slope periods (April 1-August 15 and August 16-October 31, respectively) and the almost constant winter period (November 1-March 31). Multiple models are applied to 4-week estimation periods in order to forecast the fifth ones. No model seems to adequately fit the daily water demand distribution, due to its noisy fluctuation; however, the two slope periods are much easier to fit, and therefore forecast, than the winter period, as it is indicated by the averages of the R square values of the 4-week period samples for the 3 different periods (Table 3). One explanation is that the touristic water demand, which does not exhibit weekly seasonality and mainly causes the ascent and descent adversarial the peak, is massively over-scaling the daily fluctuation of water demand. This implies that, with more micro-scaling recording of water demand, it seems possible to achieve an almost real-time water demand forecasting in the summer period, when it is more essential.

Table 3: Adequacy of multiple models' fit to daily water demand, through the average R square for the 3 periods

	April 1-August 15	August 16-October 31	November 1-March 31
Average R square through the 4-week period samples	0,569	0,687	0,306

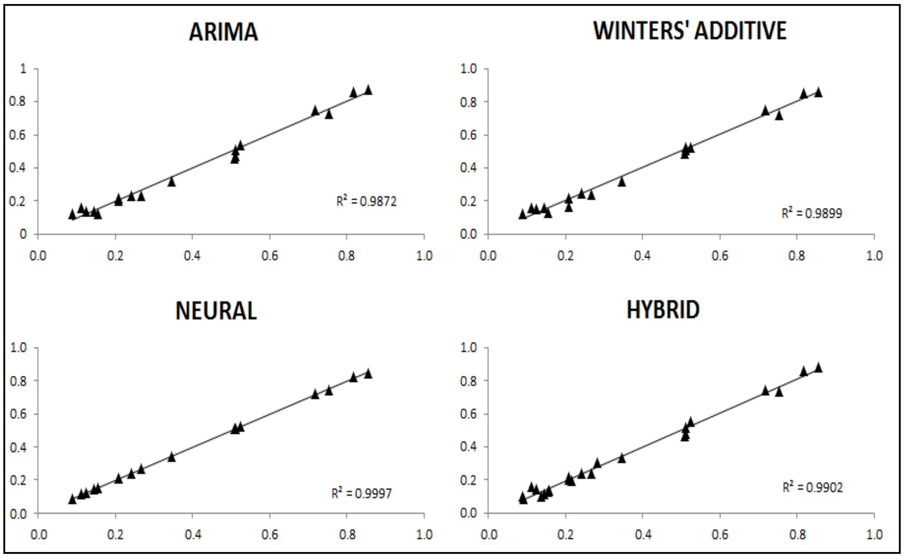

Fig. 3. Scatter plots of actual and simulated monthly water demand for the ARIMA, Winters' Additive, Neural and Hybrid models

4. Conclusions

In this paper, we interpreted the urban water demand patterns through a three-year time series for Skiathos island, a representative Greek touristic resort. The summer demand peak is six times multiple the winter level and demand patterns roughly consist of an ascending spring-summer slope, a descending summer-autumn slope and an almost constant winter part. We exploited the strengths of different popular approaches for urban water demand forecasting via univariate time series analysis. The effort was made for quarterly, monthly and daily values of demand. The only applicable model among the ones we applied for the quarterly averaged data that gave meaningful values was the Winters' additive. For the monthly-averaged data, the first approach is devoted to statistical time series modeling using ARIMA, the second one to Winters' Additive Exponential Smoothing, the third one to ANNs and the fourth to a hybrid approach. Each one of the investigated approaches presents the advantageous characteristics of linear and nonlinear modeling. We presented the forecast produced for 2014 for the quarterly and monthly averaged data. The daily time series could not be described by any model, due to the specific nature of the observed values, which were not water meter readings representing actual water consumption, but were groundwater-pumping quantities that were stored in a tank for future urban use. However, one could conclude that it is easier to simulate increasing and decreasing summer demand, rather than the winter demand, which is relatively low. Four common measures of accuracy were applied to assess models performance. Performance of individual time series models was compared to decide the best model so as to ensure appropriate simulation and forecast of water demand time series.

Acknowledgements

This work was supported by the project ISS EWATUS—Integrated Support System for Efficient Water Usage and Resources Management—which is implemented in the framework of the EU 7th Framework Programme, Specific programme Cooperation Information and Communication Technologies; Grant Agreement Number 619228.

References

[1] C.S. Laspidou, S.E. Gialis, A. Loukas, and A. Kungolos, Analysis and evaluation of the operational status of municipal wastewater treatment plants in the Dodecanese prefecture in Greece, Water Policy 13 (2011) 287-297.
[2] S. Alvisi, M. Franchini, and A. Marineli, A short-term pattern-based model for water demand forecasting. J. of Hydroinformatics 9(1) (2007) 39-50
[3] B.R. Billings and V.C. Jones, Forecasting Urban Water Demand. American Water Works Associations, second ed., AWWA, 2008, pp. 7.
[4] T.G. Mamo, I. Juran, and I. Shahrour, Urban Water Demand Forecasting Using the Stochastic Nature of Short Term Historical Water Demand and supply Pattern, J. of Water Resource and Hydraulic Engineering 2(3) (2013) 92-10.
[5] G.E.P. Box, , G.M. Jenkins, Time Series Analysis Forecasting And Control. Holden-Day, San Francisco, USA, 1976.
[6] K.W. Chau, A review on integration of artificial intelligence into water quality modeling, Marine Pollution Bulletin 52 (2006) 726–733.
[7] R. Rojas, Neural Networks: A Systematic Introduction. Springer-Verlag, Berlin, 1996.
[8] D. Machiwal and M. Kumar Jha, Hydrologic Time Series Analysis:Theory and Practice, Springer, 2012.
[9] E.A. Donkor, T.A. and Mazzuchi, Urban Water Demand Forecasting: A Review of Methods and Models, J. of Water Resources Planning and Management 140(2) (2012) 146-159.
[10] C.T. Su, L.I. Tong, C.M. Leou, Combination of time series and neural network for reliability forecasting modeling, J. of Chinese Industrial Engineering 14 (1997) 419–429.
[11] P.G. Zhang, Time series forecasting using a hybrid ARIMA and neural network model, Neurocomputing 50 (2003) 159–175.
[12] D. O. Faruk, A hybrid neural network and ARIMA model for water quality time series prediction, Engineering Applications of Artificial Intelligence 23 (2010) 586–594.
[13] L. Jentgen, H. Kidder, R. Hill and S. Conrad, Energy management strategies use short-term water consumption forecasting to minimize cost of pumping operations, J. American Water Works Association 99(6) (2007) 86–94.
[14] G. Zhang, E.B. Patuwo, M.Y. Hu, Forecasting with artificial neural networks: the state of the art, Int.J. Forecasting 14 (1998) 35–62.
[15] E.D. McKenzie, General exponential smoothing and the equivalent ARMA process, J. Forecasting 3 (1984) 333–344.
[16] G.E.P. Box, G. Jenkins, Time Series Analysis, Forecasting and Control, Holden-Day, San Francisco, CA, 1970.
[17] S.A. DeLurgio, Forecasting Principles and Applications. Tom Casson, New York, 1998.
[18] G.E.P. Box, G.M. Jenkins, G.C. Reinsel, Time Series Analysis, Forecasting and Control, Prentice Hall, Englewood Cliffs, NJ, USA, 1991.
[19] C. Dougherty, Introduction to Econometrics, Oxford University Press, 1992
[20] P. R. Winters,. Forecasting sales by exponentially weighted moving averages, Management Science 6 (1960) 324-342.
[21] H. Allende, C. Moraga, R. Salas, Artificial neural networks in time series forecasting: a comparative analysis, Kybernetik 38 (2002) 685–707.
[22] D. Rumelhart, J. McClelland, Parallel Distributed Processing, MIT Press, Cambridge, MA, 1986.
[23] P. Coulibaly, F. Anctil, B. Bobee, Hydrological forecasting with artificial neural networks: the state of the art, Canadian Journal of Civil Engineering 26 (1999) 293–304.